

Sampling Survey Design Presidential Election Quick Count Sumatera Island

Wardhani Utami Dewi^{1*}, Warsono², Khoirin Nisa³

^{1*,2,3}Department of Mathematics, Universitas Lampung, Lampung, Indonesia

*corresponding author: dewiutamiwardhani@gmail.com

Received December 22, 2022; Received in revised form December 28, 2022; Accepted January 1, 2023

Abstract. The number of TPS on the island of Sumatra is very large, in order to save time and money in conducting surveys, a sampling survey design was created. The purpose of this study is to predict the results of the presidential election on the island of Sumatra. The TPS sample frame was obtained in four stages where each stage used a sampling technique, namely the first and second stages used stratified random sampling, the third stage used systematic random sampling, and the last used clusters. The results obtained are with different TPS sample sizes showing the same results. The victory in the presidential election on the island of Sumatra was won by candidate pair number two. Then compared with the overall TPS population in Sumatra. Based on the population, the second candidate pair is also superior. So it can be concluded that the use of a survey sampling design in this study is appropriate in predicting the results of the elected president election.

Keywords: Random sampling, TPS, Pemilu



This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

INTRODUCTION

The voting process in general elections (Pemilu) is carried out simultaneously throughout Indonesia. Elections conducted in the territory of Indonesia are very broad and many areas are difficult to reach, causing the voting process to take a long time and costs a lot. After the voting is completed, the next process is vote counting at all polling stations in Indonesia. This will definitely take quite a long time. With this phenomenon, the quick calculation technique is used as a new alternative to obtain temporary results from the elections that have been held. This new method is commonly called the Quick Count (Veronika et al., 2020; Williams & Curiel, 2020).

In statistics, quick count is a topic worthy of discussion because in the calculation process it uses the sampling method (Mendoza & Nieto-Barajas, 2016; Ridwan & Aji, 2021). Quick count is able to provide fast information, low cost, and accurate results. This can be realized if the sampling is carried out correctly and the recording procedure is carried out correctly, even though only using a TPS sample, the quick count results can describe the election results well (Ulya et al., 2018).

According to Rahman et al. (2022) Several sampling techniques were used for the quick count, namely simple random sampling, stratified random sampling, systematic random

sampling, and cluster random sampling. The sampling method has its own criteria, determining sample size, sampling technique, up to the level of accuracy of each method. Penelitian sebelumnya oleh Jian and Lai (2021) mengombinasikan teknologi big data secara organik dengan survei pengambilan sampel, data yang diperoleh bisa lebih kaya dan efektif. Kemudian dilakukan penelitian berupa simulasi penentuan sampel quick count Pemilu Legislatif di Kota Bengkulu (Gumilar et al., 2022).

The presidential and vice presidential elections which took place on April 17, 2019 were elections which were held simultaneously in 34 provinces. One of the areas that has an impact on the election results is the island of Sumatera, because it is the third largest island in Indonesia. The island of Sumatera is spread over 10 provinces, and it is recorded that there are 171,028 polling stations. Therefore, the authors are interested in conducting a quick count by making a survey design to obtain a TPS sample frame on the island of Sumatera. This report aims to predict quickly (quick count) the results of the vote count in the 2019 presidential and vice presidential elections on the island of Sumatera.

METHOD

Data Sources and Research Variables

The data used in this study is secondary data obtained from the official website of the 2019 Indonesian presidential election KPU (KPU, 2019). The variable used in this research is the number of votes for each presidential candidate pair in each polling station on the island of Sumatera in the 2019 Indonesian presidential election. The total polling stations on the island of Sumatera are 171,028, spread across 10 provinces.

Analysis Method

There are 4 stages in carrying out a quick count, in which each stage is carried out by taking samples starting from the province until the sample frame is collected using different sampling methods, the following is an explanation of each stage in the quick count:

Stage 1

The population used is all provinces on the island of Sumatera. Sampling was carried out using stratified random sampling method. The strata are formed based on the territorial division of the island of Sumatera. Then the sample size taken from each stratum is calculated proportionally.

Stage 2

After obtaining provinces from each stratum described in stage 1, then the selected provinces are made into strata. Each selected province was taken as a district/city sample using a stratified random sampling method. The sample size taken from each stratum is also calculated proportionally.

Stage 3

Districts/cities have been selected, enter stage 3, namely sampling in the form of districts using a systematic random sampling method.

Stage 4

In this 4th stage, samples will be taken in the form of TPS from each sub-district. In this case the selected sub-districts were taken from the sub-districts using the cluster random sampling method. If the TPS sample has been selected, it will be calculated with a different

sample size, to compare that the two samples selected are able to represent the population and the characteristics of the election results for the presidential election on the island of Sumatera.

RESULT AND DISCUSSIONS

The stages of the quick count method in this study are as follows:

Stage 1:

In this stage 1, a sample of several provinces will be taken from the population elements of the provinces on the island of Sumatera. Sumatera Island consists of 10 provinces, namely Aceh Province, North Sumatera Province, West Sumatera Province, Riau Province, Riau Islands Province, Jambi Province, Bengkulu Province, South Sumatera Province, Bangka Belitung Province and Lampung Province.

It is expected that in taking the sample it can represent the characteristics of all elements of a heterogeneous population, so that it uses a stratified random sampling method (Ward, 2022; Zahid et al., 2022). Consider a finite population sized N units with L homogeneous subgroups called strata (Rabee et al., 2021). The design of this survey is to group members of the population into certain groups that do not overlap, then take them randomly from each group. Therefore, so that each part of the region is represented, the island of Sumatera will be divided geographically into 3 regions, namely the northern, central and southern parts as shown in the following figure.



Figure 1. Regional division of Sumatera Island

This will simplify as well as save time and cost in taking provincial samples. The steps for taking samples using stratified random sampling are as follows:

1. Define strata
The strata determined in this study are the northern, central, and southern regions. The division of the region has homogeneity in terms of geographical area.
2. Placing each member of the population in the appropriate strata

Table 1. Geographical Strata Division

Level	Province
North (Strata 1)	aceh North Sumatera
Middle (Strata 2)	North Sumatera West Sumatera Riau Riau islands Jambi
South (Strata 3)	Bengkulu South Sumatera Bangka Belitung Lampung

3. Determine the number of samples taken from each stratum
 The author determines the sample size (n) to be taken is 3 Provinces. To be proportional. Then, the samples taken from each stratum are calculated using the following formula:

$$n = \frac{\text{Group Population (Stratum)}}{\text{Total Population}} \times \text{specified number of samples} \tag{1}$$

$$\text{Strata 1} = \frac{2}{10} \times 3 = 0.6 \text{ (rounded to 1)}$$

$$\text{Strata 2} = \frac{4}{10} \times 3 = 1.2 \text{ (rounded to 1)}$$

$$\text{Strata 3} = \frac{4}{10} \times 3 = 1.2 \text{ (rounded to 1)}$$

Based on the calculation above, it was found that each stratum was taken randomly as many as 1 sample province.

4. Sampling from each stratum with the Simple Random Sampling method
 Each stratum is taken 1 random sample using simple random sampling because each stratum is homogeneous. Sampling used the help of Microsoft Excel, so the selected samples were Aceh Province, West Sumatera Province and Lampung Province. The following is an illustration of the survey design in diagrammatic form:

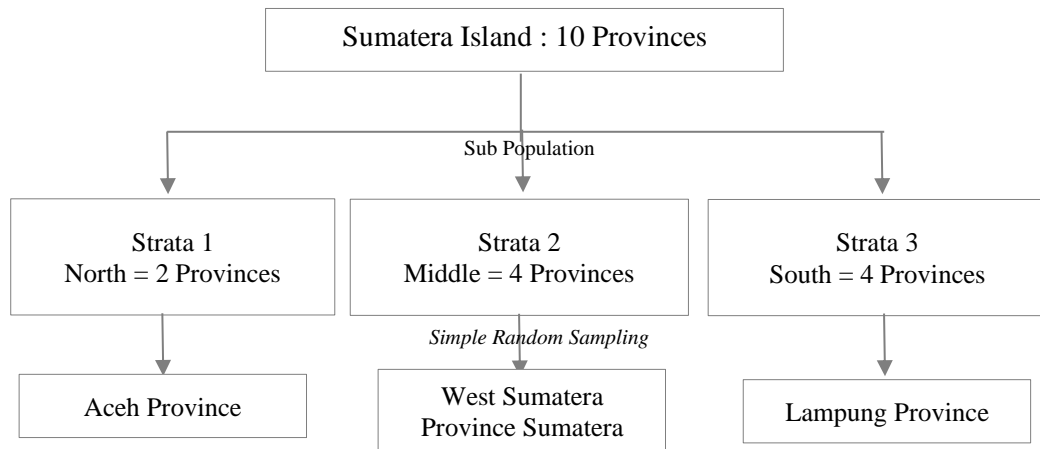


Figure 2. Stratified random sampling diagram for regions in Sumatera

Stage 2:

After the three Provinces were selected, the next step was to take samples in the form of Regencies/Cities from each Province that had been selected in the previous stage. The sampling method used is stratified random sampling. The reason for using this sampling method is that the formed strata are homogeneous in certain characteristics, including having the same ethnicity or customs, several political parties supporting the same, and government workflows that tend to be the same. The steps in taking district/city samples using the stratified random sampling method are as follows (Garg & Pachori, 2022):

1. Define strata
The author determines Aceh Province, West Sumatera Province, and Lampung Province as strata. These three Provinces have been selected using the sampling method described in step 1.
2. Placing each member of the population in the appropriate strata
Each stratum consists of several geographically appropriate members, namely Regencies/Cities, which are presented in the following table:

Table 2. Distribution of Provincial Strata

Level	Regency/City
Aceh Province (Strata 1)	Aceh Barat
	Aceh Barat Daya
	⋮
	Pidie Jaya
	Simeulue
West Sumatera Province (Strata 2)	Agam
	Dharmasraya
	⋮
	Solok Selatan
	Tanah Datar

Lampung Province (Strata 3)	Kota Bandar Lampung Kota Metro : Tulang Bawang Barat Way Kanan
-----------------------------	--

3. Determine the number of samples taken from each stratum
 The author determines the sample size (n) to be taken is 6 districts/cities. So that the proportional sample taken from each stratum is calculated by the following formula (2):
 Strata 1 = $\frac{23}{57} \times 6 = 2.42$ (rounded to 2)

Strata 2 = $\frac{19}{57} \times 6 = 2$

Grade 3 = $\frac{15}{57} \times 6 = 1.83$ (rounded to 2)

Based on the calculation above, it was found that for strata 1 2 samples were taken, 2 samples were taken for stratum 2, and 2 samples were also taken for stratum 3.

4. Sampling from each stratum with the Simple Random Sampling method
 For each stratum, n samples are taken randomly using simple random sampling because the population is homogeneous. Sampling using the help of Microsoft Excel. The results obtained were that 6 districts/cities were selected, namely Southeast Aceh, North Aceh, Sijunjung, Solok City, Tanggamus, and South Lampung. The following is an illustration of the survey design in diagrammatic form:

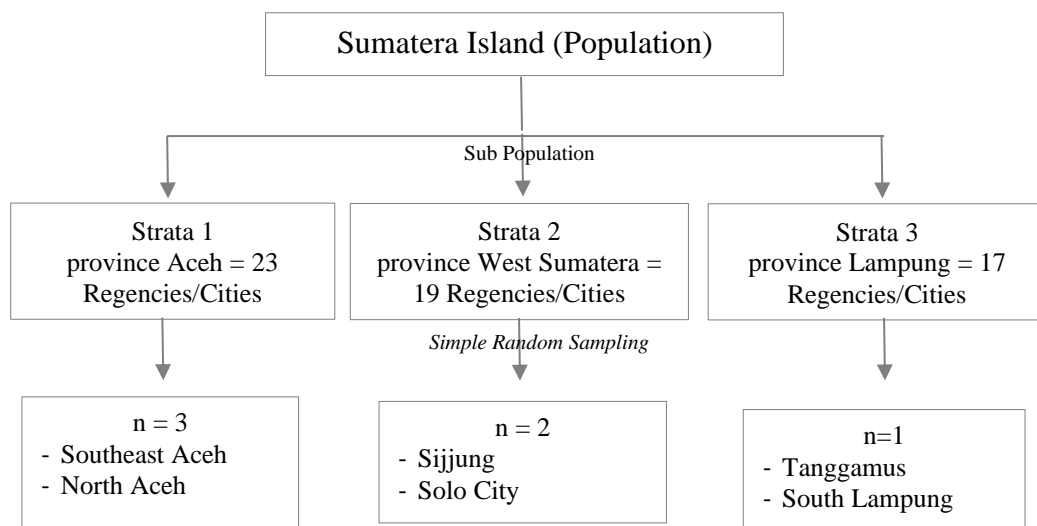


Figure 3. Province stratified random sampling diagram

Stage 3:

Continue to stage 3, at this stage samples will be taken in the form of sub-districts using the systematic random sampling method (Rahman et al., 2022; Wright et al., 2015). The reason for using this method is that the population is quite large. In addition, it will facilitate

sampling, because the sub-district data is coherent based on regional mapping. The following is the data for the selected districts/cities along with their sub-districts.

Table 3. Distribution of districts/municipalities

No	Regency/City	Subdistrict
1.	Southeast Aceh	Babul Makmur Babul Rahmah : Semadam Tanoh Alas
2.	North Aceh	Baktiya Baktiya Barat : Tanah Luas Tanah Pasir
3.	Sijunjung	IV Nagari Kamang Baru Koto Vii : Sumpur Kudus Tanjung Gadang
4.	Solo City	Lubuk Sikarah Tanjung Harapan
5.	Tanggamus	Air Naningan Bandar Negeri Semuong : Ulu Belu Wonosobo
6.	South Lampung	Bakauheni Candipuro : Way Panji Way Sulan

Sampling steps using systematic random sampling method (Baltes & Ralph, 2022):

1. Determine the sample size (n) to be taken from all members of the population (N). The author determines that the sample size to be taken is 8 district samples from 89 members of the population.
2. Determine the number of groups
To determine the number of groups, use the following formula:
 $k = \frac{89}{10} = 8.9$ (rounded to 9), meaning there are 10 groups and every 9th unit is taken.
3. Randomly assign serial numbers to the 89 sub-districts from 1, 2, 3 to 89. These numbers are sorted based on the sub-district order that has been recorded in the 2019 election data.

Dividing all members of the population into 10 groups. Then the first group (group A) will be obtained containing districts with serial numbers 1 to 10, the second group (group B) with serial numbers 11 to 20, and so on until group J.

- Then take one unit sample at random in group A (first). The author uses Microsoft Excel to take a random sample in group A and number 3 which is selected. After that, sampling is carried out in the next group for sample units that are in line (have the same distance) with sample number 3. So the sampling was carried out completely randomly only on the first sampling from the first group. After the first sample is taken, the second, third and so on samples are taken systematically from the second, third and so on groups. Members of the population that are sampled in this study are members of the population who have the following numbers:

Table 4. Sampling technique with systematic random sampling

Group	A	B	C	D	E	F	G	H	I	J
No. Selected	3	12	21	30	39	48	57	66	75	84

- So that the number chosen shows the sample in the form of sub-districts, as shown in the following table:

Table 5. Selected sub-district samples

Selected Number	Subdistrict
3	Babussalam
12	Lawe Sigala-Gala
21	Dewantara
30	Nibong
39	Syamtalira Aron
48	Sijunjung
57	Gisting
66	Pugung
75	Jati Agung
84	Sidomulyo

Stage 4:

At this stage 4, a sample frame will be taken in the form of a collection of TPS taken from each kelurahan. Sampling based on population elements grouped according to adjacent locations (Ekan & Bala, 2017). The sampling method at this stage is cluster random sampling. The population used is 10 districts selected in stage 3, so that there are 10 clusters with heterogeneous characteristics and geographically the elements are widely spread . Within each cluster there are village elements.

So that we can know that the survey design that has been made represents the characteristics of the population on the island of Sumatera, the sample size is made into two. The first sample size was taken as many as 4 clusters, and the second was taken 5 clusters. So that the number of sample frames in the form of TPS will also be different. The steps for taking samples using cluster random sampling (Ma & Song, 2021).

Sample size of 4 clusters

1. Determine the required sample

Researchers took a sample of 4 sub-districts from 10 existing sub-districts.

2. Determine the elements that are in each cluster

In this case the elements of each cluster are kelurahan which are geographically heterogeneous. The number of elements in each cluster will be presented in the following table.

Table 6. Number of sub-districts from each selected sub-district

No	Subdistrict	Number of Villages
1	Babussalam	27
2	Lawe Sigala-Gala	35
3	Dewantara	15
4	Nibong	20
5	Syamtalira Aron	35
6	Sijunjung	9
7	Gisting	9
8	Pugung	27
9	Jati Agung	21
10	Sidomulyo	16

3. Choose clusters randomly

As many as 4 clusters out of 10 clusters will be selected using simple random sampling with the help of Microsoft Excel, in order to obtain 4 clusters (sub-districts) including the Districts of Syamtalira Aron, Dewantara, Lawe Sigala-gala, Sidomulyo, and Gisting. The following is an illustration of the cluster random sampling survey design:

Cluster 1 27 Wards	Cluster 2 35 Wards	Cluster 3 15 Wards	Cluster 4 20 Wards	Cluster 5 35 Wards
Cluster 6 9 Wards	Cluster 7 9 Wards	Cluster 8 27 Wards	Cluster 9 21 Wards	Cluster 10 16 Wards

4 clusters were taken as a sample using simple random sampling

Cluster 1 27 Wards	Cluster 2 35 Wards	Cluster 3 15 Wards	Cluster 4 20 Wards	Cluster 5 35 Wards
Cluster 6 9 Wards	Cluster 7 9 Wards	Cluster 8 27 Wards	Cluster 9 21 Wards	Cluster 10 16 Wards

Figure 4. Cluster random sampling with a sample size of 4 clusters

So the results obtained are as follows:

Table 7. Obtained results based on selected TPS samples (4 clusters)

Cluster (District)	Number of Villages	Number of TPS	Paslon 1	Paslon 2
5 (Syamtalira Aron)	35	63	726	9435
3 (Dewantara)	15	67	6514	4631
2 (Lawe Sigalagala)	35	131	1502	23080
10 (Sidomulyo)	16	200	23929	11878
Total	101	461	32671	49024
		Percentage	39.99%	60.01%

Sample size of 5 clusters

Using the same steps as before, only the sample size is 5 clusters. So when illustrated using a picture is as follows:

Cluster 1 27 Wards	Cluster 2 35 Wards	Cluster 3 15 Wards	Cluster 4 20 Wards	Cluster 5 35 Wards
Cluster 6 9 Wards	Cluster 7 9 Wards	Cluster 8 27 Wards	Cluster 9 21 Wards	Cluster 10 16 Wards

A sample of 5 clusters was taken using simple random sampling

Cluster 1 27 Wards	Cluster 2 35 Wards	Cluster 3 15 Wards	Cluster 4 20 Wards	Cluster 5 35 Wards
Cluster 6 9 Wards	Cluster 7 9 Wards	Cluster 8 27 Wards	Cluster 9 21 Wards	Cluster 10 16 Wards

Figure 4. Cluster random sampling with a sample size of 5 clusters

So the results obtained are as follows:

Table 8. Obtained results based on selected TPS samples (5 clusters)

Cluster (District)	Number of Villages	Number of TPS	Paslon 1	Paslon 2
5 (Syamtalira Aron)	35	63	726	9435
3 (Dewantara)	15	67	6514	4631
2 (Lawe Sigalagala)	35	131	1502	23080
10 (Sidomulyo)	16	200	23929	11878
7 (Gisting)	9	137	14497	9758
Total	110	598	47618	58782
		Percentage	44.51%	55.48%

The total TPS on the island of Sumatera is 171,028 TPS. It would be very inefficient and would definitely cost a lot of money to count the votes as a whole. So that this research will carry out a quick count process with several sampling methods to minimize time and costs in the vote counting process. The author divides the sampling into 4 stages with different methods, this is done according to the needs and characteristics of the areas on the island of Sumatera. Starting from the election of the province until obtaining a TPS sample using the sampling method. The author divides 2 different sample sizes, so that it can be seen that the survey design made represents the entire population on the island of Sumatera. The first sample size was taken as many as 461 TPS, resulting in a vote for Paslon 1 of 32,671 or 39.99% and a total vote for Paslon 2, namely 49,024 or 60.01%. This means that for a sample size of 461 polling stations, the presidential and vice presidential candidates, namely Paslon 2, are superior to Paslon 1 .

Then if we look at it based on the second sample size, which is 598 TPS, the vote acquisition for Paslon 1 is 47,618 or 44.51% and for the presidential candidate Paslon 2 it is 58,782 or 55.48%. In a sense, Paslon 2's presidential and vice presidential candidates got more votes. The use of two different sample sizes gives the same results, namely the prediction of the victory of the presidential and vice presidential candidates owned by Paslon 2 for Sumatera Island.

Based on data obtained from the 2019 election website, out of a total of 1171,028 polling stations on the island of Sumatera, it shows that the total vote acquisition for Paslon 1 was 13,226,504 or 42.70%. As for the Paslon 2 candidate pair, there were 17,746,070 or 57.29%. Based on these data, if we compare the total vote acquisition from all polling stations on Sumatera Island with the selected sample polling stations, it shows that the vote acquisition is not much different, where the Paslon 2 pair is predicted to win for the Sumatera Island region. So it can be concluded that the use of the sampling method at each stage until the TPS sample is obtained is appropriate and able to represent the election population on the island of Sumatera.

CONCLUSION

This study each stage in the survey design used a different sampling method, adjusting to the characteristics and geography of the area. The sample frame taken is the selected TPS sample. The survey design was made in 4 stages with each method using stratified random sampling, stratified random sampling, systematic random sampling, and cluster random sampling. The author determines two TPS sample sizes, so that they can be compared with TPS populations on the island of Sumatra. The results obtained with the first sample size of 461 TPS and the second sample size of 598 TPS show that the Paslon 2 candidate pair is superior to Paslon 1. When viewed on a population basis, the results of calculations for all polling stations on the island of Sumatra also show that the Paslon 2 pair is superior.

This means that the use of survey sampling at every stage up to the vote count based on the TPS sample is correct. What's more, it is more efficient to use even with a minimal sample size and at the same time saves on survey implementation costs. namely the results of calculations for all polling stations on the island of Sumatra also show that the Paslon 2 pair is superior. This means that the use of survey sampling at every stage up to the vote count based on the TPS sample is correct. What's more, it is more efficient to use even with a minimal sample size and at the same time saves on survey implementation costs. namely the results of calculations for all polling stations on the island of Sumatra also show that the Paslon 2 pair is superior. This means that the use of survey sampling at every stage up to the vote count based on the TPS sample is correct. What's more, it is more efficient to use even with a minimal sample size and at the same time saves on survey implementation costs.

REFERENCE

- Baltes, S., & Ralph, P. (2022). Sampling in software engineering research: A critical review and guidelines. *Empirical Software Engineering*, 27(4), 1–31. <https://doi.org/10.1007/s10664-021-10072-8>
- Etikan, I., & Bala, K. (2017). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, 5(6), 149. <https://doi.org/10.15406/bbij.2017.05.00149>
- Garg, N., & Pachori, M. (2022). Log Type Calibration Estimator of Population Mean in Stratified Sampling. *Journal of the Indian Society for Probability and Statistics*, 1–27. <https://doi.org/10.1007/s41096-022-00123-8>
- Gumilar, A. T., Nugroho, S., & Keraman, B. (2022). Simulation of Sample Determination Quick Count Legislative Elections In Bengkulu City. *Journal of Statistics and Data Science*, 1(1), 30–37. <https://doi.org/10.33369/jsds.v1i1.21012>
- Jian, X., & Lai, W. (2021). Application of Statistical Sampling Survey in the Context of Big Data Era : Based on China's 1% population sample survey in 2015. *2021 International Conference on Big Data Analysis and Computer Science (BDACS)*, 22–25. <https://doi.org/10.1109/BDACS53596.2021.00013>
- KPU. (2019). *Hasil Hitung Suara Pemilu Presiden & Wakil Presiden Ri 2019*.
- Ma, T., & Song, F. (2021). A trajectory privacy protection method based on random sampling differential privacy. *ISPRS International Journal of Geo-Information*, 10(7), 454. <https://doi.org/10.3390/ijgi10070454>
- Mendoza, M., & Nieto-Barajas, L. E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. *Electoral Studies*, 43, 124–132. <https://doi.org/10.1016/j.electstud.2016.06.007>
- Rabee, S., Hamed, R., Kassem, R., & Rashwaan, M. (2021). A Goal Programming Approach for Multivariate Calibration Weights Estimation in Stratified Random Sampling.

- Mathematics and Statistics*, 9(3), 326–334. <https://doi.org/10.13189/ms.2021.090314>
- Rahman, M. M., Tabash, M. I., Salazadeh, A., Abduli, S., & Rahaman, M. S. (2022). Sampling techniques (probability) for quantitative social science researchers: a conceptual guidelines with examples. *Seeu Review*, 17(1), 42–51. <https://doi.org/10.2478/seeur-2022-0023>
- Ridwan, Y., & Aji, R. H. S. (2021). Determine Sample Size for Precision Results on Quick Count. *Proceedings of The International Conference on Data Science and Official Statistics*, 2021(1), 810–823. <https://doi.org/10.34123/icdsos.v2021i1.121>
- Ulya, S. F., Sukestiyarno, Y., & Hendikawati, P. (2018). Analisis Prediksi Quick Count Dengan Metode Stratified Random Sampling Dan Estimasi Confidence Interval Menggunakan Metode Maksimum Likelihood. *Unnes Journal of Mathematics*, 7(1), 108–119.
- Veronika, R., Hayati, M. N., & Purnamasari, I. (2020). Metode Quick Count dan Analisis Autokorelasi Spasial Menggunakan Indeks Moran (Studi Kasus: Pemilihan Presiden Indonesia Tahun 2019 di Kalimantan Timur). *Statistika*, 8(2). <https://doi.org/10.26714/jsunimus.8.2.2020.121-126>
- Ward, B. (2022). Resolving the Raven Paradox: Simple Random Sampling, Stratified Random Sampling, and Inference to Best Explanation. *Philosophy of Science*, 89(2), 360–377. <https://doi.org/10.1017/psa.2021.26>
- Williams, J., & Curiel, J. (2020). Analysis of the Quick Count in the 2019 Bolivia Election. *Washington, DC, Center for Economic and Policy Research*.
- Wright, A. I., Grabsch, H. I., & Treanor, D. E. (2015). RandomSpot: A web-based tool for systematic random sampling of virtual slides. *Journal of Pathology Informatics*, 6(1), 8. <https://doi.org/10.4103/2153-3539.151906>
- Zahid, E., Shabbir, J., & Alamri, O. A. (2022). A generalized class of estimators for sensitive variable in the presence of measurement error and non-response under stratified random sampling. *Journal of King Saud University-Science*, 34(2), 101741. <https://doi.org/10.1016/j.jksus.2021.101741>